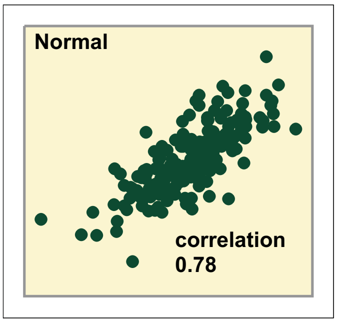
A common misconception with correlated lognormals

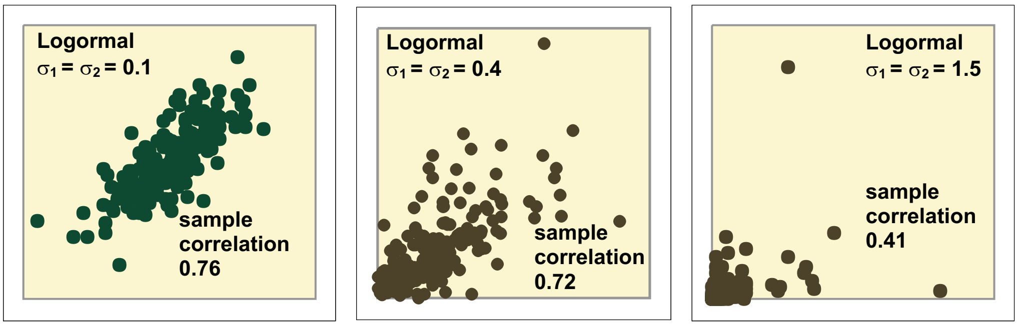
Actuaries frequently need to find covariances or correlations between variables such as when finding the variance of a sum of forecasts (for example in P&C reserving, when combining territories or lines of business, or computing the benefit from diversification).

Correlated normal random variables are well understood. The usual multivariate distribution used for analysis of related normals is the multivariate normal, where correlated variables are linearly related. In this circumstance, the usual linear correlation (the *Pearson correlation*) makes sense.
See paper on **Correlation and Linearity** for more details. Click here

However, when dealing with lognormal random variables (whose logs are normally distributed), if the underlying normal variables are linearly correlated, then the correlation of lognormals changes as the variance parameters change, even though the correlation of the underlying normal does not.

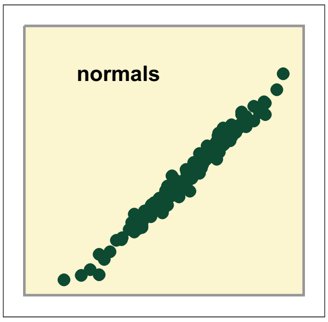


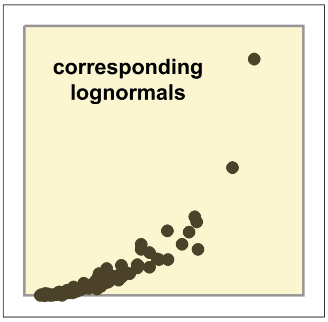
All three lognormals below are based on normal variables with correlation 0.78, as shown left, but with different standard deviations.



We cannot measure the correlation on the log-scale and apply that correlation directly to the dollar scale, because the correlation is not the same on that scale.

Additionally, if the relationship is linear on the log scale (the normal variables are multivariate normal) the relationship is no longer linear on the original scale, so the correlation is no longer *linear* correlation. The relationship between the variables in general becomes a curve:





It may not be a good idea to directly measure a linear correlation between variables that are not linearly related, even when we want to find the linear correlation between them. It will not be fully using the available information, and the loss of information becomes worse as the variables become less like a normal. If the logs of the lognormal variables (that is, the corresponding normal random variables) are linearly (or approximately linearly) related, we can best measure correlation on the log scale.

If we then need the linear correlation of lognormal variables, and believe that the logs of the random variables are multivariate normal, we should measure the correlation on the log scale, but then calculate the corresponding correlation on the original scale.

What about when the logs are not linearly related?

Some practitioners use copulas to model more general kinds of relationship between variables than that implied by a linear correlation. In general, lognormals related by a non-Gaussian copula will still not be linearly related (indeed, apart from a few special cases, they can't be made linearly related). Information from the copula used to model the relationship should be used to derive the correlation on the log scale. In cases where that is difficult to do exactly, Taylor series expansions (the so-called *delta method*) can simplify the calculations and yet often yield a very good approximation.

For further study on **Correlation and Linearity**
click here